DOCUMENT RESUME

ED 398 239 TM 024 739

AUTHOR Cross, Lawrence H.
TITLE Grading Students. ERIC/AE Digest.
INSTITUTION ERIC Clearinghouse on Assessment and Evaluation, Washington, DC.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO EDO-TM-95-5
PUB DATE Oct 95
CONTRACT RR93002002
NOTE 4p.; Adapted from "Testing Memo 6: What Kinds of Grades Should Be Averaged" and "Testing Memo 11: Absolute versus Relative Grading Standards: What Does a Percentage Mean," Office of Measurement and Research Services, Virginia Polytechnic Institute and State University, Blacksburg, VA 24060.
AVAILABLE FROM ERIC Clearinghouse on Assessment and Evaluation, The Catholic University of America, Department of Education, O'Boyle Hall, Washington, DC 20064 (free).
PUB TYPE Information Analyses - ERIC Clearinghouse Products (071)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; Elementary Secondary Education; *Grades (Scholastic); *Grading; Scores; *Scoring; *Student Evaluation
IDENTIFIERS Composite Scores; *Standardized Scores; T Values; Variability; *Z Scores

ABSTRACT

Third Digest addresses several aspects of assigning grades, beginning with a discussion of the variability of test scores, and reviewing the use of standardized scores, ideas on assigning letter grades, and recommendations for grading. An important characteristic of grades as initially recorded, the variability of the scores of each test or assignment, is seldom questioned. The best approach to dealing with variability is to record and average standardized test scores. To calculate standardized scores, it is necessary to know the standard deviation of the scores prior to standardization. Deriving the standard score, commonly called a z-score, is explained, as is transforming z-scores into T-scores for convenient calculation. If T-scores are computed for every test, averaging them will provide a composite score from which the influence of the variability of the scores has been eliminated. The distribution of T-scores should be inspected in assigning letter grades, but the assignment across a range of average scores is essentially an arbitrary matter of professional judgement. Some specific recommendations are made for passing scores and grading. (Contains 5 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 398 239

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

GRADING STUDENTS

ERIC/AE Digest Series EDO-TM-95-5
The Catholic University of America

Lawrence H. Cross

TM024739

BEST COPY AVAILABLE

CUA

ERIC/AE DIGEST

DEPARTMENT OF EDUCATION THE CATHOLIC UNIVERSITY OF AMERICA

EDO-TM-95-5 October 1995

Grading Students

Lawrence H. Cross
Virginia Polytechnic Institute and State University

Some instructors record letter grades for tests and assignments, and others record numerical values, often the percent correct on tests. Later, under either method, the grades are averaged, often employing a weighting process designed to make some grades count more heavily than others. Discussion of the merits of different approaches usually centers around the question of whether it is better to average letter or numerical grades or around some feature of the weighting process.

This Digest discusses several aspects of assigning grades. First, an issue that underlies for both approaches, variability of test scores, is discussed. The use of standardized scores is presented as a solution to the variability problem, ideas on assigning letter grades, and recommendations are then presented.

Variability

An important characteristic of the grades as initially recorded is seldom questioned, namely the variability of the scores of each test or assignment. Indeed, it is ironic that instructors may go to considerable trouble to weight grades according to their perceived importance when, in fact, the result may be quite different from what was intended, due to failure to account for differences in score variability from one test to another.

To see how this outcome could occur, consider a course in which the midterm examination was much more difficult than the instructor intended; scores ranged from 35% to 95% with an average of 65%. Further assume the instructor did not view this outcome as desirable and, with the intention of being fair to the students, included a large proportion of easy questions on the final examination. Their presence caused a great reduction in score variation. Final examination scores ranged only from 88% to 100% with and average of 94%. Only a small number of harder questions kept everyone from earning very high scores in a narrow range. The result was that differences from one student to another in final course averages were largely attributable to scores on the midterm. Thus, a student's achievement in the latter part of the course was effectively devalued, which was hardly fair or in keeping with the presumed intention that grades reflect achievement across the entire course.

Standardized scores

The best approach to avoiding situations like the one just presented is to record and average standardized test scores. In order to calculate standardized scores it is necessary to know the standard deviation of the scores prior to standardization. This statistic is a measure of how "spread out" the scores are and is explained in any elementary statistics text. Though nearly any program reporting test results will include this statistic, a fair approximation for most classroom tests may be obtained by subtracting the lowest score from the highest and dividing by 4. In the example above, the standard deviations are about 15 and 3 percentage points respectively for the midterm and final examinations. A standard score is then the number of standard deviations the number-right or percentage score is above or below average. Commonly called a z-score, its formula is: $z = (x - \bar{x})/s$, in which x is the observed score, $\bar{x}$ is the average score, and s is the standard deviation. Then scores of 80% and 97% on the midterm and final respectively would each yield z-scores of 1.0, because both are one standard deviation above average. Similarly, scores of 50% and 91% would correspond to z-scores of -1.0.

It may be difficult to work with z-scores, because half of them will be negative and all will probably lie between -3 and 3. Therefore, it is convenient to transform the z-scores into T-scores as follows: $T = 50 + 10z$. T-scores will have a mean (average) of 50 and a standard deviation of 10. Thus, a T-score of 60 represents a number-right score one standard deviation above the average. If the distribution of scores approximates the shape of the normal curve, about 16% of the T-scores will be above 60 and about 10% above 63. Similarly, about 16% of T-scores will be below 40 and about 10% below 37.

If T-scores are computed for every test, averaging them will provide a composite score from which the influence of the variability of the scores has been eliminated. (Strictly speaking, if more than two scores are to be averaged, the intercorrelations among the scores should be taken into consideration in order to control for the degree of "overlap." However, simple averaging of T-scores should produce a good approximation of the more precise result.) T-scores are typically provided for multiple-choice tests processed by measurement services offices at universities. Moreover, T-scores can be calculated for any numerically evaluated non-test assignments you may wish to include in the course composite. Like other scores, T-scores may be weighted differentially. For example, if you wish to weight the final exam twice as much as the midterm, multiply the T-scores from the final by 2, add the midterm T-scores and divide by 3.

TM024739

ERIC Clearinghouse on Assessment and Evaluation

The Catholic University of America, Department of Education, O'Boyle Hall, Washington, DC 20064 202 319-5120

It should be noted at this point that T-scores report only a student's relative position in the class and not an absolute measure of achievement. However, we contend that the difficulty level of nearly all academic tests is arbitrary and that, regardless of the scoring method, they provide nothing more than ranking information. The concern of this Digest is that the scores be averaged in a manner consistent with the instructor's intention.

Assigning letter grades

Finally, when the T-scores have been averaged, there is the problem of assigning letter grades for the course. Until this point, we have been able to speak with conviction, deducing conclusions logically through arguments based on statistical principles. However, when it is necessary to determine the dividing line between As and Bs or Ds and Fs, there is no such clear-cut approach available. Of course, if student X's average is higher than student Y's, student Y's letter grade must not be higher than student X's, but beyond this recommendation our best advice is to inspect the distribution of average T-scores with the following questions in mind:

- What is a typical letter grade distribution for a course of this type with this kind of student?
- Are there any circumstances which might warrant altering this "typical" distribution, e.g., did the course progress especially well or poorly?
- Where in the distribution are key students whose work you know especially well, students you believe might deserve especially good or poor grades for reasons other than test performance?
- Where are naturally occurring "breaks" in the distribution of average T-scores? (There is no "scientific" reason for letting these points determine letter grades, but if their use is not inconsistent with other considerations, it will help to prevent hard feelings on the part of students who otherwise might miss a better grade by one T-score point.)

Two ideas to be avoided or at least questioned in determining letter grades are:

- That the T-score spread should be the same for each letter grade.
- That an equal number of As and Fs, Bs and Ds, should necessarily be awarded.

Finally, it must be remembered that assignment of letter grades a cross a range of average scores is essentially arbitrary and a matter of professional judgement.

Recommendations:

1. Do not prescribe percent-correct score ranges for letter grades in your course syllabus. Instead, indicate that you will exercise your professional judgment as to what constitutes A, B, C, etc., only after reviewing the test scores. You may wish to indicate only tentative letter grades for any given test and base final test grades on the average of standardized scores from several tests.

2. When testing higher level cognitive skills, vary the difficulty of the questions so as to discriminate among all skill levels. Include items sufficiently difficult to challenge even the most talented students and a few items sufficiently easy that most will answer correctly. The latter may include a disproportionate number of lower level cognitive skills. The average percent-correct score should be somewhere in the range between 50% and 70% in order to maximize discrimination among achievement levels.

3. Alert students to the fact that the test may be more difficult than what they are accustomed to, but that the percent-correct scores will be interpreted in a relative, rather than an absolute sense.

4. Determine the minimum passing score on each test by identifying items that you (and/or your colleagues) judge to represent essential knowledge, or that you (or they) believe should be correctly answered by any student deserving of a passing grade. Base the passing score on the percentage of the total points these questions contribute. You may wish to compute a separate score for items so identified, but it is probably sufficient to use this percentage regardless of which items are answered correctly.

5. Use your professional judgment to determine cut points between grades. You might consider the test performance of students about whom you have independent knowledge of achievement via other assignments or previous courses. Naturally occurring breaks in the score distribution can suggest cut points between letter grades that might minimize the number of students clamoring at your door for the extra point or two needed for the next higher grade. Ultimately, judgements in this area are subjective and should be acknowledged as such when implemented.

6. Do not feel obliged to "grade on the curve" whereby a specified percentage of students will receive each letter grade. To do so can be as arbitrary and capricious as to adopt prescribed percentage ranges for assigning scores, especially in smaller classes.

This Digest was adapted with permission from *Testing Memo 6: What kinds of grades should be averaged,* and *Testing Memo 11: Absolute versus relative grading standards: What does a percentage mean*, Office of Measurement and Research Services, Virginia Polytechnic Institute and State University, Blacksburg, VA 24060

Further Reading

Airasian, P. (1994) *Classroom Assessment*, Second Edition, NY: McGraw-Hill.

Lysne, A. (1984) Grading of Student's Attainment: Purposes and Functions *Scandinavian Journal of Educational Research*, 28(3), 149-65.

Nottingham, M. (April, 1988) Grading Practices--Watching Out for Land Mines. *NASSP Bulletin*, 72 (507), 24-28.

Ornstein, A.C.(April, 1989) The Nature of Grading. *Clearing House*, 62 (8), 365-69.

Terwilliger, J.S (1989) Classroom Standard Setting and Grading Practices. *Educational Measurement: Issues and Practice*, 8(2), 15-19.

BEST COPY AVAILABLE

This publication was prepared with funding from the Office of Educational Research and Improvement, U.S. Department of Education, under contract RR93002002. The opinions expressed in this report do not necessarily reflect the positions or policies of OERI or the U.S. Department of Education. **Permission is granted to copy and distribute this ERIC/AE Digest.**